Claims

The claims are:

1. A computer system for classifying items, comprising:

a computer system component that applies probabilistic dependency models, one for each of a plurality of categories, to an item to provide with respect to each of the plurality of categories an indication of whether the item belongs;

wherein the probabilistic dependency models collectively employ outputs from a plurality of classifiers; and

the outputs employed by the probabilistic dependency models vary among the probabilistic dependency models.

2. The computer system of claim 1, wherein the dependency models collectively employ one or more reliability indicators.

3. The computer system of claim 1, wherein the probabilistic dependency models are decision trees.

4. The computer system of claim 1, wherein the items are texts.

5. A computer system for classifying items, comprising:

a computer system component that applies a probabilistic dependency model to classify an item, wherein the probabilistic dependency model contains dependencies on one or more classical outputs from one or more classifiers and dependencies on one or more reliability indicators.

6. The computer system of claim 5, wherein the computer system outputs a quantitative measure relating to confidence that the item belongs in a category.

7.   The computer system of claim 6, wherein the probabilistic dependency models are decision trees.

8.   The computer system of claim 6, wherein the items are texts.

9.   A computer system, comprising:

a first computer system component that learns, from training examples, probabilistic dependency models for classifying items according to one or more reliability indicators together with classical outputs from one or more classifiers.

10.   The computer system of claim 9, further comprising a second computer system component that repeatedly invokes the first component to learn probabilistic dependency models employing various potentially effective reliability indicators and compares the performances of the resulting probabilistic dependency models to identify reliability indicators that are effective.

The computer system of claim 10, wherein the second component automatically selects the potentially effective reliability indicators.

11.   The computer system of claim 9, wherein the first computer system component treats the classical outputs from classifiers and the reliability indicators in the same manner.

12.   The computer system of claim 9, wherein the probabilistic dependency models are decision trees.

13.   The computer system of claim 9, wherein the items are texts.

14.   A computer readable medium having computer executable instructions for performing steps comprising:

implementing a plurality of classifiers adapted to receive and classify at least one item, the plurality of classifiers each generating a score related to classification of the at least one item; and

for each of one or more categories, facilitating classification, selection, and/or utilization of the at least one item with a probabilistic dependency model that employs one or more of the scores and, in addition, one or more reliability indicators.

15.   The computer readable medium of claim 14, wherein:

the instructions implement a different probabilistic dependency model for each of two or more categories;

the probabilistic dependency models are based on subsets of parameters selected from the group consisting of the scores and the reliability indicators; and

the subsets of parameters vary among the probabilistic dependency models.

16.   A system for classifying items, comprising:

means for determining a model that classifies the items based on a probabilistic approach that combines information about the items including one or more classical outputs of classifiers and one or more attributes of the items other than classical outputs of classifiers; and

means for applying the model to classify the items.

17.   A computer-readable medium having stored thereon a data structure useful in classifying items, comprising:

first data fields containing data representing an attribute to test, wherein the attributes represented include both classical classifier outputs and reliability indicators;

second data fields corresponding to the first data fields and containing data representing values against which to compare the attributes;

third data fields containing data representing classifier outcomes;

fourth data fields facilitating determination of relationships among instances of the first, second, and third data fields, the relationships having a decision tree structure with the first and second data fields corresponding to decision nodes and the third data fields corresponding to leaf nodes.

18.   The computer-readable medium of claim 13, wherein the data represented by the first data fields comprises classical classifier outputs from a plurality of classifiers.

19.   A method of generating a classifier, comprising:

obtaining a set of training examples;

applying a probabilistic approach that uses the training examples to develop a model that combines evidence to provide an output relating to whether an item belongs in a category; and

storing the model in a computer-readable media for use as a classifier;

wherein the evidence comprises one or more classical outputs of other classifiers and one or more attributes of the item other than classical outputs of classifiers.

20.   A method of identifying useful reliability indicators, comprising

obtaining potentially useful reliability indicators;

applying the method of claim 19 using various of the potentially useful reliability indicators as evidence; and

comparing the resulting classifiers to identify which of the potentially useful reliability indicators are, in fact, useful.

21.   The method of claim 19, wherein the model is a decision tree.

22.   The method of claim 19, wherein the evidence comprises classical outputs from two or more classifiers.

23. A method of classifying items, comprising:

obtaining the items in computer readable format,

employing a computer to classify the item using a classifier generated according to the method of claim 19.

The method of claim 23, wherein the items are texts.

24. A method of classifying items, comprising:

applying probabilistic dependency models, one for each of a plurality of categories, to an item stored in computer readable format to provide an output relating to whether the item belongs in the category with respect to each of the plurality of categories;

wherein the probabilistic dependency models collectively contain dependencies on outputs from a plurality of classifiers; and

the outputs considered by the probabilistic dependency models vary among the probabilistic dependency models.

25. The method of claim 24, wherein the dependency models collectively contain dependencies based on one or more reliability indicators.

26. The method of claim 24, wherein the probabilistic dependency models are decision trees.

27. A method of combining a plurality of classifiers to classify items, comprising:

sequentially applying tests to the items to obtain test results; and

classifying the items based on the test results.

wherein the sequence of tests applied varies among the items in that the outcome of one or more tests affects whether another test is applied, whereby the classifiers utilized vary depending on the items.

28. The method of claim 27, wherein one or more of the tests involves a reliability indicator.

29. The method of claim 27, wherein the classifiers are applied to the items selectively to avoid applications of classifiers that will not be utilized.